DOCUMENT RESUME

ED 260 125 TM 850 452

AUTHOR Lockwood, Robert E.
TITLE The Alabama High School Graduation Examination Experience: Technical Concerns.
PUB DATE Apr 85
NOTE 41p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, IL, April 1-4, 1985). For a related document, see TM 850 451.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Standards; *Cutting Scores; Elementary Secondary Education; *Equated Scores; Evaluation Criteria; Evaluators; *Graduation Requirements; High School Graduates; Item Analysis; Mastery Tests; *Minimum Competency Testing; *Scoring; State Departments of Education; State Programs; *State Standards; Student Certification; Test Construction; Testing Programs; Test Theory
IDENTIFIERS Alabama; *Alabama High School Graduation Examination

ABSTRACT

A technical advisory committee assisted the Alabama State Department of Education in establishing passing scores and equating test forms of the Alabama High School Graduation Examination (AHSGE). The committee examined four methods of setting cutting scores: (1) Jensen's theoretical method--probability theory is used to set standards that are independent of both test characteristics and examinee performance; (2) judgmental--a modification of Angoff's method in which judges estimate the percentage of minimally competent examinees who will respond correctly to each item; (3) empirical--Livingston and Zieky's contrasting groups method--the passing score is the one at which half of the students are masters; and (4) empirical--Livingston and Zieky's borderline groups method, in which the average performance of borderline students becomes the passing score. Reading, mathematics, and language test results were analyzed using each of the four approaches. The final recommended passing scores were an average of the standards set with the borderline group approach and Angoff's judgmental method: 79 percent correct in reading, 65 percent in mathematics, and 77 percent in language. The committee also used detailed procedures to pre-equate the AHSGE to assure identical item formats, content, and difficulty among test forms. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED260125

The Alabama High School Graduation Examination Experience: Technical Concerns

Robert E. Lockwood

ALABAMA STATE DEPARTMENT OF EDUCATION

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. E Lockwood

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented in symposium -- The Alabama High School Graduation Examination Experience -- at the meeting of the National Council on Measurement in Education, Chicago, April 1985.

TM 850 452

The Alabama High School Graduation Experience: Technical Concerns

There are many technical and quasi-technical issues that must be addressed during the development and implementation of a high school graduation examination. In Alabama a technical advisory committee, composed of educators and measurement specialists from within the state, was formed to assist the State Department of Education (SDE) with these issues as they pertained to the Alabama High School Graduation Examination (AHSGE). This paper describes two of the issues addressed by the technical advisory committee, namely the methods used to establish passing scores and procedures for equating test forms. Consequently, the procedures described herein are the result of the work of the members of the committee and the SDE and not solely the author.

STANDARD SETTING

A number of approaches for setting standards of acceptable performance have been proposed (cf. Angoff, 1971; Berk, 1976; Ebel, 1979; Nedelsky, 1954). The various approaches have been reviewed and discussed (Glass, 1978; Hambleton, 1978; Jaeger, 1976; Meskauskas, 1976; Millman, 1973; Shepherd, 1980), and publications detailing steps to be followed in using the more promising ones have appeared (Livingston & Zieky, 1982; Popham, 1978).

Given all of the approaches to standard setting, persons who

must establish standards to separate those who pass, for example, from those who do not pass have a dilemma. If similar standards resulted from the application of the different methods, the problem would be alleviated. Unfortunately, such is not always the case. Studies (cf. Andrew & Hecht, 1976; Brennan & Lockwood, 1979; Halpin, Sigmon, & Halpin, 1983; Koffler, 1980; Skakun & Kling, 1980) reveal that standards set with the various methods are, more often than not, different.

After considering a number of approaches to standard setting and reviewing comparative studies which have been conducted, the technical advisory committee decided to consider standard setting procedures from within three major frameworks: theoretical, judgmental, and empirical.

Representing theoretical approaches was a method proposed by Jensen (1978). With this method, probability theory is used to set standards that are independent of both test characteristics and examinee performance.

The judgmental approach selected was a modification of Angoff's method (1971). Using this method involves analyses and decisions regarding the items comprising the tests for which passing scores are to be set.

Empirical methods included the contrasting groups and borderline group approaches. With each of these approaches, students' mastery status and test performance are considered in

establishing passing scores.

A more detailed description of each of the standard-setting methods used and the passing scores that resulted are presented in the following sections.

The Theoretical Approach

In the theoretical approach as proposed by Jensen (1978), probability theory is used to differentiate between test scores resulting from knowledge of test content and test scores resulting from random guessing. The passing score, determined from the number of items on the test and the number of options per item, is the minimum score for which there is a specified probability that the score was not obtained by random guessing.

Discussing his approach, Jensen said:

> If a perfectly internally consistent, n-item, minimum competency test is scored on the number of right answers basis, the following will be true:
>
> 1. The possible test scores are zero and n.
> 2. Each obtained test score is a true knowledge (T) score.
> 3. The minimum passing point (lowest test score which separates the "know nothings" from the "know-it-alls") is the lowest test score greater than zero; Min pp = 1.
> 4. The maximum passing point equals n and is identical in effect to the Min pp, regardless of whether n is 1 or 1000.

If a fallible n-item test with an average of k choices per item is administered to a large group and then is scored on the number of right answers basis, the following will be true:

1. Fifty percent of those with a T score of zero (50% of those randomly marking answers to all items) will obtain a test score equal to or greater than n/k.
2. One percent of those with a T score of zero (1% of those randomly marking aswers to all items) will obtain a test score equal to or greater than $np + 2.326\sqrt{npq}$ (where $p = 1/k$ and $q = 1 - p$). (p. 1)

Determining the passing score as $np + 2.326\sqrt{npq}$ is appropriate in low-risk situations where the consequences of allowing less-than-competent examinees to pass is relatively minor. In situations of significant risk, the passing score is raised until there is a 50% proability that an examinee with an obtained score has a true score equal to or greater than the scores of 99% of those examinees who randomly marked answers to all the items. In high-risk situations where those not competent should not pass, the passing score is raised until there is a 99% chance that the examinee's true score is above the maximum score of 99% of those randomly guessing. This shift in the passing scores set in the low-risk, significant risk, and high-risk situations are graphically depicted in Figure 1.

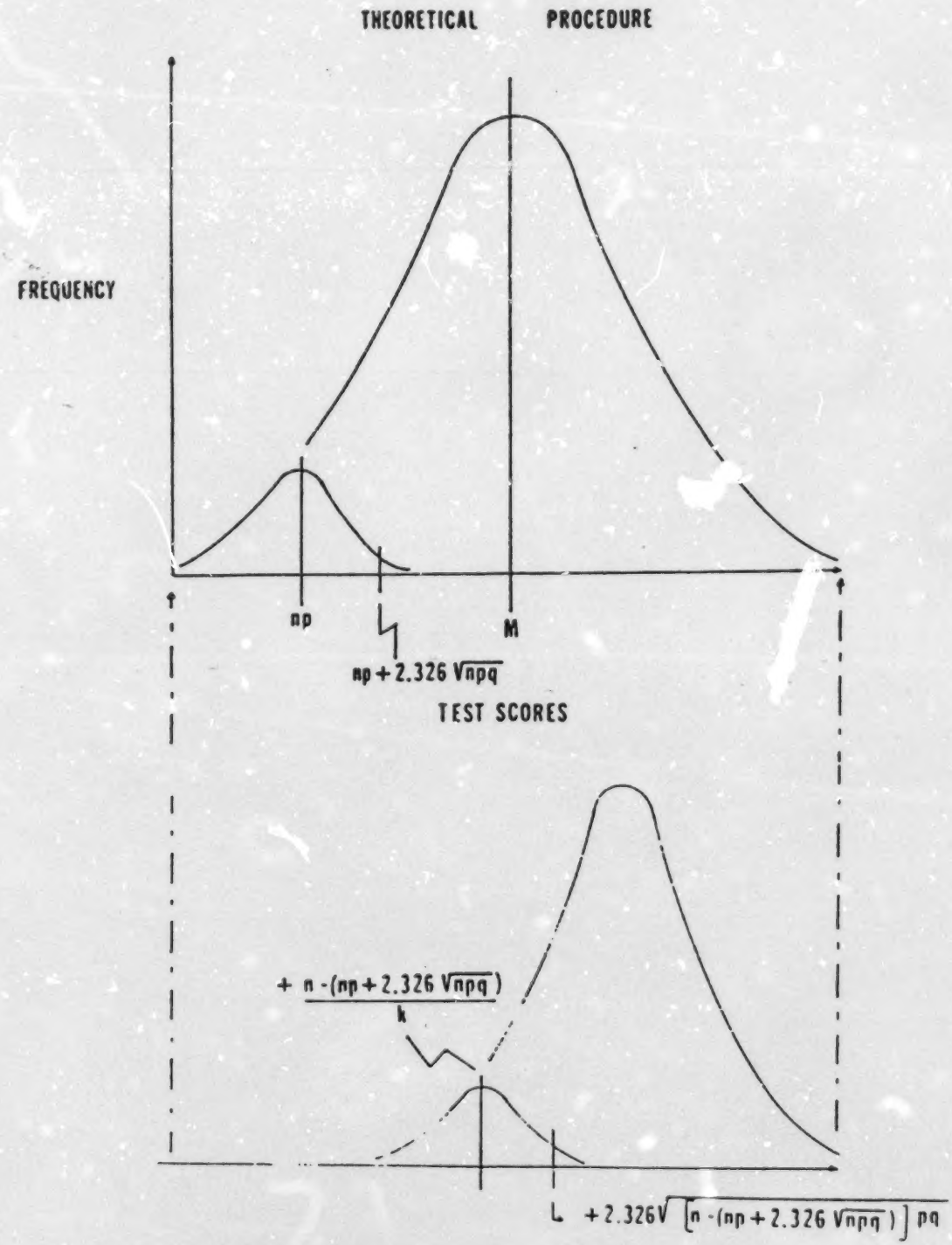


Figure 1. Theoretically derived passing scores.

From "The Theoretical Approach: Method and Results" by Robert E. Lockwood. Paper presented in symposium, A Multimethod Approach for Establishing the Passing Score on the Alabama High School Graduation Examination, conducted at the meeting of the National Council on Measurement in Education, New Orleans, April 1984.

Passing the AHSGE was considered a high-risk situation. This conclusion was predicated on a sound rationale. The purpose of the graduation examination is to guarantee the credibility of the high school diploma--students receiving it should be certifiably competent in designated skills. Setting the theoretical passing scores at a high-risk level served to ensure that examinees passing the graduation tests would indeed be competent, particularly since the theoretical standards were to be benchmarks below which the recommended passing scores could not fall. Even though the theoretical standards were high, students initially failing the AHSGE are given up to three additional opportunities to take the examination, and they should be able to pass if they have indeed mastered the competencies.

Results from the application of the Jensen method of standard setting are given in Table 1.

Table 1

Results of the Theoretical Standard Setting Procedure

Subject Area	Number Items	Options Per Item	Passing Score and Percentage
Reading	80	4	49 (61%)
Mathematics	95	4	57 (60%)
Language	119	4	70 (59%)

The passing scores of 49 on the reading test, 57 on the mathematics test, and 70 on the language test became the minimum passing scores considered.

The Judgmental Procedure

The judgmental procedure selected was one based on an approach recommended by Angoff (1971). With his procedure, judges establish passing scores by estimating the percentage of a hypothetical group of a minimally competent examinees they think will respond correctly to each item on the test. The sum of these percentages is the minimally acceptable score for each judge, and the average of all scores for the judges becomes the passing score.

With the Angoff method, the selection of the judges is therefore critical. It is extremely important that they have the background and expertise to enable them to render expert judgments. After much discussion, the technical advisory committee decided to ask Alabama classroom teachers with either experience teaching remedial high school classes or teaching general classes in grades nine and ten to serve as judges. Eleven mathematics teachers and 23 English teachers were chosen. The group of English teachers was subdivided so that 11 judged the reading test of the AHSGE while 12 judged the language test. The mathematics group judged the math test.

All judges were brought together for training and to establish the passing scores. A first step was the establishment of an

agreed upon definition of minimally competent. It was:

> A student who can complete the required Carnegie units and has a minimum knowledge of [reading, language, or mathematics] to graduate from high school.

Judges were then given the list of the competencies in their respective area that are required for high school graduation and test items measuring these competencies. They were asked to read the competencies and corresponding test items carefully. Considering the minimally competent students as had previously been defined, the judges were then asked to estimate the number of items, out of 100 such as those they had reviewed, that the minimally competent should pass.

Working independently, judges completed their item estimations for each of the competencies in their area of expertise. They were then given average item difficulties for the items measuring each competency which had been determined during an earlier pilot testing of the AHSGE. After being given an explanation of these item difficulties, the judges were asked to review their earlier judgments and make modifications if they felt any were needed. Decisions made during this second round of judging became the ones used in the determination of the passing scores with the judgmental procedure.

Percentages (number out of 100) of the items the minimally competent should pass were averaged across judges for each of the

competencies in reading, mathematics, and language. This respective average was multiplied by the number of items measuring each competency to give the expected number of items per competency the minimally competent student should pass. The sums of these products became the passing scores set with the judgmental method for the reading, mathematics, and language tests. Results can be seen in Tables 2, 3, and 4.

The Empirical Method

Empirical approaches considered were the contrasting groups and borderline group methods (Livingston & Zieky, 1982).

With the borderline group method, individuals (for example, teachers) familiar with the competence of students in the areas being tested are asked to identify those students who are at or near the borderline with respect to mastery of the competencies of interest. The average performance of these identified borderline students on the test becomes the passing score.

In the contrasting groups approach, judges (e.g., teachers) choose those students clearly thought to be masters or nonmasters of the competencies in question. Any one of several methods may then be used to calculate passing scores using actual test data for these students. Selected was the method suggested by Livingston and Zieky (1982) wherein the passing score is the one at which 50% of the students are masters with the assumption being that at the

Table 2

Passing Score Set for Reading Using the Judgmental Procedure

	Competency	Number of Items Measuring Competency	Average Percentage Expected by Judges	Number of Items Based on Expected Percentage
I-1, 2; IV-13	Associate selected written words and phrases which occur in daily life situations with their literal meanings; recognize and use various forms and application blanks used in everyday life.	8	76.36	6.1088
II-1	Determine word meaning using specific context clues	5	72.73	3.6365
III-1	Determine word meanings from a knowledge of prefixes and suffixes.	4	68.64	2.7456
III-2	Recognize common abbreviations and symbols.	4	83.18	3.3272
IV-1	Use alphabetical organization to locate information needed.	6	76.36	4.5816
IV-3	Select appropriate reference sources in a given situation.	4	79.09	3.1636
IV-4	Know the uses of the different sections of a book.	4	73.18	2.9272
IV-5	Know the uses of the different sections of the newspaper.	4	80.45	3.2180
IV-6	Determine the main idea in a passage.	5	70.00	3.5000
IV-7	Find details in a written passage.	5	76.36	3.8180
IV-8	Use maps, graphs, charts, and tables.	6	74.09	4.4454
IV-9, 10, 11	Determine main idea and detail from selected materials on government, law, money management, health, and nutrition.	9	70.91	6.3819

BEST COPY

Table 2 continued

	Competency	Number of Items Measuring Competency	Average Percentage Expected by Judges	Number of Items Based on Expected Percentage
IV-14	Read for information from materials such as advertisements, yellow pages, labels, directories, and catalogs.	6	81.55	4.8930
V-1	Follow directions and instructions.	6	79.05	4.7454
VI-1	Distinguish fact from fiction.	4	71.36	2.8544
Total		80	75.43	60.3466

Table 3

Passing Score Set for Mathematics Using the Judgmental Procedure

	Competency	Number of Items Measuring Competency	Average Percentage Expected by Judges	Number of Items Based on Expected Percentage
I-1	Read and write numerals and word names for numbers	3	76.00	2.2800
I-2	Compare decimals; compare fractions with unlike denominators	2	57.73	1.1546
I-3	Add, subtract, multiply, and divide whole numbers.	6	82.73	4.9638
I-4	Add, subtract, multiply, and divide decimals.	7	70.09	4.9063
I-5	Add, subtract, and multiply fractions.	6	61.82	3.7092
I-6	Change percents to decimals and decimals to percents.	3	65.55	1.9665
I-7	Determine discounts, down payments, commissions, taxes, ratios, and percents	4	59.27	2.3708
I-8	Convert fractions to decimals and decimals to fractions.	4	58.36	2.3344
I-9	Find averages of whole numbers and decimal numbers.	3	69.18	2.0754
I-10	Change units of measure.	4	51.18	2.0472
II-1	Solve problems involving time.	3	63.18	1.8954
II-2	Measure distance by use of a ruler.	4	60.00	2.4000
II-3	Measure capacities.	3	65.55	1.9665
II-4	Find perimeters.	4	69.09	2.7636
II-5	Find rectangular areas.	4	53.64	2.1456
III-1	Read and interpret graphs.	2	63.18	1.2636

Table 3 continued

	Competency	Number of Items Measuring Competency	Average Percentage Expected by Judges	Number of Items Based on Expected Percentage
III-2	Read and interpret tables and charts.	2	64.27	1.2854
III-3	Interpret scale drawings.	3	62.27	1.8681
IV-1	Make approximations by rounding numbers.	3	66.00	1.9800
V-1, 3	Recognize simple plane geometric figures, parallel and perpendicular lines.	4	77.73	3.1092
VI-1	Determine change.	2	70.91	1.4182
VI-2	Use comparative shopping skills	2	63.91	1.2782
VI-3	Compare the cost of cash and credit purchases.	2	57.73	1.1546
VI-4	Allocate time and money.	3	61.91	1.8573
VI-5	Complete checks and money orders.	3	67.55	2.0265
VI-6	Solve rate, time, distance problems.	4	64.64	1.9392
VI-7	Compute salaries and wages.	4	61.36	2.4544
Total		95	65.29	62.0230

BEST COPY

Table 4

Passing Score Set for Language Using the Judgmental Procedure

	Competency	Number of Items Measuring Competency	Average Percentage Expected by Judges	Number of Items Based on Expected Percentage
I-1	Observe pronoun-antecedent agreement	9	83.33	7.4997
I-2	Use correct forms of nouns and verbs.	15	80.25	12.4997
II-1	List ideas in sequential order.	5	79.33	3.9665
II-2	Arrange sentences into a logically organized paragraph.	9	75.58	6.8022
II-3	Arrange written directions in logical order.	4	81.92	3.2768
IV-1	Include in a message or request all necessary information (who, what, when, where, how, or why).	4	82.50	3.3000
IV-2	Determine what information is missing from a message, an announcement, or a process explanation; or what information is irrelevant.	4	82.75	3.3100
V-1	Identify proper format of a business letter.	9	77.75	6.9975
V-2	Identify the proper format of a friendly letter (social correspondence).	8	80.92	6.4736
VI-1	Complete a common form such as a driver's license application, change of address form.	9	83.75	7.5375
VII-1	Identify words frequently used in daily activities.	9	86.25	7.7625
VIII-1, 2, 3, 4	Identify question marks, periods, and exclamation points to punctuate sentences.	16	81.50	13.0400
VIII-5	Identify quotation marks to set off a direct quotation.	4	73.33	2.9332

BEST COPY

	Competency	Number of Items Measuring Competency	Average Percentage Expected by Judges	Number of Items Based on Expected Percentage
VIII-6	Identify the colon after "the following" and "as follows," the salutation of a business letter, and the hour in numerals.	4	75.00	3.0000
IX-1	Identify the capitalization of the first word in every sentence, the pronoun I, and proper nouns.	10		
Total		119	81.03	96.4290

Note. Tables 2, 3, and 4 are from "The Judgmental Approach: Method and Results" by Rodney W. Roth. Paper presented in symposium, A Multimethod Approach for Establishing the Passing Score on the Alabama High School Graduation Examination, conducted at the meeting of the National Council on Measurement in Education, New Orleans, April 1984.

lower test score levels the proportion of unqualified examinees who pass increases.

In setting passing scores with the empirical methods, schools throughout Alabama were first ranked according to average performance by their students on the California Achievement Tests. From each of five low-to high-achievement strata identified, schools were randomly selected, and, within each of those schools, from 6 to 88 11th grade students were randomly selected to take the AHSGE.

In each school, the teacher or teachers most familiar with each student's respective mastery of the basic competencies being measured by the reading, mathematics, and language tests were asked:

> "In your opinion, will this student have a sufficient knowledge of [reading, mathematics, language] to graduate from high school?"
>
> Possible answers: "Yes," "It is questionable.", or "No."

This information was collected prior to testing, generally from students' mathematics teachers for mathematics and English teachers for language and reading, and coded on the student answer sheets after testing.

Students for whom the answer was "yes" were designated as masters for standard setting purposes. Those for whom the answer was "no" became the nonmasters. The borderline group was made up

of the students for whom the answer was "It is questionable."

In setting the contrasting groups passing scores using the method described by Livingston and Zieky, test score intervals were established, and the smoothed percent of masters and nonmasters at each interval was determined using the moving averages method of smoothing. Results for reading, mathematics, and language are shown in Tables 5, 6, and 7.

For the reading, mathematics, and language tests the decision rule was to set the passing scores at the lower limit of the score interval where 50% of the examinees fell into the mastery category and 50% in the nonmastery category since it was generally deemed equally serious to misclassify either masters or nonmasters. When an exact 50-50 point was not available, the cut-off was determined by finding the interval where the proportions of students classified as masters and nonmasters were closest to the 50-50 point. The higher interval was chosen where the 50-50 point (or a close approximation) occurred twice since, due to a larger number of examinees at the higher point, the results should be more stable.

For the borderline group, means were computed for the reading, mathematics, and language tests. These means became the reading, mathematics, and langauge passing scores set with the borderline group method.

Passing scores set with the contrasting groups and borderline group appraoches are shown in Table 8.

Table 5

Summary of Performance of Mastery and Nonmastery Groups in Reading

Score intervals	Mastery frequency	Nonmastery frequency	Total frequency	Smoothed percent of mastery group	Smoothed percent of nonmastery group
77-80	515	4	519	*	*
73-76	208	6	214	97.63	2.37
69-72	100	10	110	93.26	6.74
65-68	38	9	47	85.71	14.29
61-64	24	8	32	73.53	26.47
57-60	13	10	23	61.90	38.10
53-56	2	6	8	47.73	52.27
49-52	6	7	13	40.00	60.00
45-48	2	2	4	37.50	62.50
41-44	1	6	7	35.00	65.00
37-40	4	5	9	38.10	61.90
33-36	3	2	5	50.00	50.00
29-32	1	1	2	44.44	55.56
25-28	0	2	2	12.50	87.50

Table continues

Table 5 (continued)

Score intervals	Mastery frequency	Nonmastery frequency	Total frequency	Smoothed percent of mastery group	Smoothed percent of nonmastery group
21-24	0	4	4	0.00	100.00
17-20	0	4	4	0.00	100.00
0-16	0	2	2	*	*

*The smoothing method cannot be used to estimate the smoothed percent at the lowest and the highest test score intervals.

Table 6

Summary of Performance of Mastery and Nonmastery Groups in Mathematics

Score intervals	Mastery frequency	Nonmastery frequency	Total frequency	Smoothed percent of mastery group	Smoothed percent of nonmastery group
92-95	69	2	71	*	*
88-91	155	3	158	98.01	1.99
84-87	121	2	123	97.52	2.48
80-83	118	5	123	97.18	2.82
76-79	105	3	108	96.08	3.92
72-75	71	4	75	95.08	4.92
68-71	56	5	61	93.12	6.88
64-67	49	4	53	89.24	10.76
60-63	36	8	44	88.64	11.36
56-59	32	3	35	87.50	12.50
52-55	23	2	25	82.93	17.07
48-51	13	9	22	73.13	26.87
44-47	13	7	20	57.14	42.86
40-43	6	8	14	52.27	47.73
36-39	4	6	10	43.33	56.67
32-35	3	3	6	45.00	55.00

Table continues

Table 6 (continued)

Score intervals	Mastery frequency	Nonmastery frequency	Total frequency	Smoothed percent of mastery group	Smoothed percent of nonmastery group
28-31	2	2	4	[illegible]00	60.00
24-27	1	4	5	25.00	75.00
0-23	0	3	3	*	*

*The smoothing method cannot be used to estimate the smoothed percent at the lowest and highest test score intervals.

Table 7

Summary of Performance of Mastery and Nonmastery Groups in Language

Score intervals	Mastery frequency	Nonmastery frequency	Total frequency	Smoothed percent of mastery group	Smoothed percent of nonmastery group
115-119	169	1	170	*	*
110-114	211	1	212	99.49	0.05
105-109	202	1	203	98.08	2.92
100-104	97	8	105	96.44	3.56
95-99	80	5	85	92.00	8.00
90-94	53	7	60	87.96	12.04
85-89	35	11	46	80.00	20.00
80-84	20	9	29	68.75	31.25
75-79	11	10	21	60.00	40.00
70-74	11	9	20	56.90	43.10
65-69	11	6	17	55.10	44.90
60-64	5	7	12	52.94	47.06
55-59	2	3	5	40.00	60.00
50-54	3	5	8	33.33	66.67
45-49	1	4	5	37.50	62.50
40-44	2	1	3	38.46	61.54

Table continues

Table 7 (Continued)

Score intervals	Mastery frequency	Nonmastery frequency	Total frequency	Smoothed percent of mastery group	Smoothed percent of nonmastery group
35-39	2	3	5	41.67	58.33
30-34	1	3	4	23.08	76.92
0-29	0	4	4	*	*

*The smoothing method cannot be used to estimate the smoothed percent at the lowest and highest test score intervals.

Note. Tables 5, 6, and 7 are from "The Empirical Approach: Method and Results" by Glennelle Halpin and D. Joyce Steele. Paper presented in symposium, A Multimethod Approach for Establishing the Passing Score on the Alabama High School Graduation Examination, conducted at the meeting of the National Council on Measurement in Education, New Orleans, April 1984.

Table 8

Passing Scores Set with the Empirical Methods

Subject Area	Number Items	Contrasting Groups	Borderline Group
		Passing Score and Percentage	Passing Score and Percentage
Reading	80	53 (66%)	64 (80%)
Mathematics	95	40 (42%)	61 (64%)
Language	119	60 (50%)	87 (73%)

Passing Scores for the Alabama High School Graduation Examination

Minimum passing scores set with the various standard setting methods are summarized in Table 9.

Table 9

Passing Scores for the Alabama High School Graduation Examination Set With the Theoretical, Judgmental, and Empirical Methods

Standard Setting Procedure	Reading	Mathematics	Language
	Passing Score and Percentage	Passing Score and Percentage	Passing Score and Percentage
Theoretical:			
Jensen	49 (61%	57 (60%)	70 (59%)
Judgmental:			
Angoff	60 (75%)	62 (65%)	96 (81%)
Empirical:			
Borderline	64 (80%)	61 (64%)	87 (73%)
Contrasting	53 (66%)	40 (42%)	60 (50%)

Since passing scores set with the theoretical method were considered benchmarks below which the recommended standards could not fall, they were thus not subjected to further consideration in the standard-setting process. Also, problems were found with the contrasting groups approach that resulted in standards set with it being dropped from consideration. First, since the percentage qualified was approximately 50% at two different score intervals on the reading test, there was no clear cut-off point. The higher point was tentatively chosen, but the lower point might have also been selected. Secondly, the number of students identified as nonmasters was lower on all three tests than the minimum number needed for a stable cut score (cf. Huynh, 1981; Popham, 1978).

Consequently, the technical advisory committee arrived at the recommended passing scores for the AHSGE by averaging the standards set with the Angoff and the borderline group approaches. This action was based on the assumption that the passing scores set with the two methods were equally valid. Furthermore, deviations from the average are at a minimum, and chances of error would therefore be less with the average than might be the case if either method were used alone.

Use of both the Angoff and borderline group methods took into account teacher judgments about the difficulty of the test items for minimally competent students and the performance levels of individual students. Standards which resulted were:

Reading — 63 of 80 items (79%)

Mathematics — 62 of 95 items (65%)

Langauge — 92 of 119 items (77%)

It was these passing scores that were recommended to the State Board of Education as minimum passing scores for the Alabama High School Graduation Examination.

EQUATING PROCEDURES

Whenever a testing program is implemented that has the impact on examinees that college admission, job selection or high school graduation examination's have, there must be a system to ensure the security of the tests. The most effective method of ensuring the security of the tests is to replace the test each time it is administered. By so doing, each examinee is provided with a test that is unique and thus eliminates any possible advantage to those examinees who may be taking the test for a second time or who have had unauthorized access to a previously used form of the test prior to the examination. In essence, the issue of test security is an issue of fairness to examinees. By the same token, the concept of fairness demands that an examinee's probability of success be dependent upon knowledge and independent of the form of the test that is taken. The melding of these two concerns, unique test forms and freedom from form bias, is the domain of test equating.

Angoff (1982) defines equating as "the process of developing a conversion from the system of units of one form of a test to the system of units of another form so that scores derived from the two forms 'after

conversion' will be equivalent and interchangeable" (p. 56). There are many methods available to derive an equating conversion formula but the underlying conceptual process is to adjust either the differences in item difficulties between the forms, assuming that both tests measure the same trait and that the groups of examinees taking the forms are of comparable ability; or use common items to adjust differences in the abilities of the two groups of examinees, assuming that they are random samples from the same population. In either case, difficulties arise when these procedures are employed with a high school graduation examination.

The Alabama High School Graduation Examination is administered to all eleventh-grade students during the second week in October. Students failing any portion of the AHSGE (reading, mathematics, and language) are provided three more opportunities to pass the failed portion(s) of the test; in the spring of the eleventh grade, the fall of the twelfth grade and the spring of the twelfth grade. Thus the population of examinees taking the spring tests is much smaller than during the fall and is comprised, for the most part, of students who have failed one or more portions of the test on previous attempts. The fall administration, to contrast, is comprised mainly of students attempting the AHSGE for the first time. Thus, any assumption that the abilities of these two groups is comparable is untenable at best. As a consequence, the use of an equating procedure that adjusts item difficulties based on differential performance between two similar groups of examinees is not appropriate in this case. Even the use of common or anchor items will not allow equating between these populations since different levels of performance on the

same items would be expected. The best and maybe only method of equating between the fall and spring is through test pre-equating.

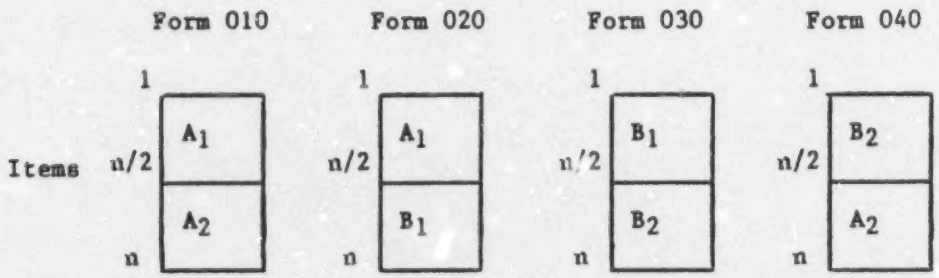
In this context test pre-equating merely refers to a system of test development that assures identical item formats and content and virtually identical item difficulties between test forms. This type of equating was employed with the Alabama High School Graduation Examination.

Without going into detail concerning the actual development and piloting of the items for the AHSGE, the developmental process resulted in two forms for each subject area of the AHSGE that were meticulously matched in terms of item content and format and which were virtually identical in item difficulties (the sum of item difficulties within .1). The item difficulties were obtained from a random sample of over 1,200 eleventh-grade students, a minimum of 600 per form, who were not required to pass the AHSGE for graduation but who were representative of future eleventh-grade students who would be required to pass the AHSGE. These pilot data were critical in forms development as well as in the standard setting process.

Because of the patience and diligence involved in developing the forms and conducting the pilot testing, it was possible to interchange, almost without exception, items measuring a particular competency on one form with those measuring that competency on the other form. It was possible, therefore, to create four equivalent forms in each subject area of the graduation examination; each form containing approximately 50% of the items from the previous form and still maintain identical predicted mean scores. A graphical presentation of the forms generated is

presented in Figure 2 where Form 010 and Form 030 are the original piloted forms.

Figure 2



This process allowed the Alabama State Department of Education to pre-equate the first four forms of the test but does not address future forms development. Regardless of the developmental strategy employed, the items would have to be pretested on a different population than the original four forms. It would be possible, however, to pretest the new items as part of a fall administration of the AHSGE. The fall administration should be used because of the ability range of the examinees and because by using the item difficulties from the administration that contained the pilot items pre-equating would be possible. This procedure may result in small changes in item difficulties from those that would have been obtained in the original pilot, but a larger problem exists. This problem revolves around the determination of the passing score for each subject area. The data from the original piloting and the original set of items were used to develop the passing standards. Any new items would have to be reviewed by the standard setting committee that developed the first standard to ensure comparability of the passing scores. Technically feasible, this process would be a logistical nightmare.

Once again the techniques of equating can be used to assure the State Board of Education that the standard for passing the AHSGE has not changed and also for assuring students that their probability of passing is independent of which form they take. After the first four forms have been administered, all future fall forms will be equated back to the original fall forms, using classical equating procedures, and the equated score will be used in the passing score formulae. In this fashion, the ability represented by the particular passing score originally established will not change and, if test development is carefully executed, the percentage of items representing that ability should not fluctuate greatly.

The final issue to be resolved is the equating of future spring forms. Since the population of students taking the spring test will always represent a different or at least more restricted range of ability, classical equating procedures are inadequate. Therefore, beginning with the third administration of the spring test, consideration is being given to using previously pre-equated forms.

References

Andrew, B. J., & Hecht, J. T. (1976). A preliminary investigation of two procedures for setting examination standards. Educational and Psychological Measurement, 36, 35-50.

Angoff, W. (1971). In R. L. Thorndike (Ed.), Educational measurement. Washington, D. C.: American Council on Education.

Angoff, W. (1982) Summary and derivation of equating methods used at ETS. In P. Holland and P. Rubin (eds) Test Equating, New York: Academic Press, 55-70.

Berk, R. A. (1976). Determination of optimal cutting scores in criterion-referenced measurement. Journal of Experimental Education, 45, 4-9.

Brennan, R. L., & Lockwood, R. E. (1979, April). A comparison of two cutting score procedures using generalizability theory. Paper presented at the meeting of the National Council on Measurement in Education, San Francisco.

Ebel, R. L. (1979). Essentials of educational measurement (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Glass, G. W. (1978). Standards and criterion. Journal of Educational Measurement, 15, 237-261.

Halpin, G., Sigmon, G., & Halpin, G. (1983). Minimum competency standards set by three divergent groups of raters using three judgmental procedures: Implications for validity. Educational and Psychological Measurement, 43, 185-196.

Halpin, G., & Steele, D. J. (1984, April). The empirical approach: Method and results. In J. E. McLean (Chair), A multimethod approach for establishing the passing score on the Alabama High School Graduation Examination. Symposium conducted at the meeting of the National Council on Measurement in Education, New Orleans.

Hambleton, R. K. (1978). On the use of cut-off scores with criterion-referenced tests in instructional settings. Journal of Educational Measurement, 15, 277-290.

Huynh, H. (1981, April). Technical and practical considerations in setting standards. Paper presented at the meetings of the American Educational Research Association and the National Council on Measurement in Education, Los Angeles.

Jaeger, R. M. (1976). Measurement consequences of selected standard-setting models. Florida Journal of Educational Research, 18, 22-27.

Jensen, O. (1978). Some logical procedures for setting passing points in the absence of direct criterion references. Unpublished manuscript, Educational Testing Service, Princeton, NJ.

Koffler, S. L. (1980). A comparison of approaches for setting proficiency standards. Journal of Educational Measurement, 17, 167-178.

Livingston, S. A., & Zieky, M. J. (1982). Passing scores: A manual for setting standards of performance on educational and occupational tests. Princeton, NJ: Educational Testing Service.

Lockwood, R. E. (1984, April). The theoretical approach: Method and results. In J. E. McLean (Chair), A multimethod appraoch for establishing the passing score on the Alabama High School Graduation Examination. Symposium conducted at the meeting of the National Council on Measurement in Education, New Orleans.

Meskauskas, J. A. (1976). Evaluation models for criterion-referenced testing: Views regarding mastery and standard setting. Review of Educational Research, 46, 133-158.

Millman, J. (1973). Passing scores and test lengths for domain-referenced testing: Views regarding mastery and standard-setting. Review of Educational Research, 43, 205-216.

Nedelsky, L. (1954). Absolute grading standards for objective tests. Educational and Psychological Measurement, 14, 3-19.

Popham, W. J. (1978). Setting performance standards. Los Angeles: Instructional Objectives Exchange.

Roth, R. W. (1984, April). The judgmental approach: Method and results. In J. E. McLean (Chair), A multimethod approach for establishing the passing score on the Alabama High School Graduation Examination. Symposium conducted at the meeting of the National Council on Measurement in Education, New Orleans.

Shepard, L. A. (1980). Standard setting issues and methods. Applied Psychological Measurement, 4, 447-467.

Skakun, E. N., & Kling, S. (1980). Comparability of methods for setting standards. Journal of Educational Measurement, 17, 229-235.